Interpreting regression models in clinical outcome studies

Freely available online OPEN ACCESS

BJR

EDITORIAL

Interpreting regression models in clinical outcome studies

D. F. Hamilton, M. Ghert, A. H. R. W. Simpson

From The British Editorial Society of Bone and Joint Surgery, London, United Kingdom

Measuring the outcome of an intervention is central to the practice of evidence based medicine, and most research papers evaluating patient outcomes now incorporate some form of patient-based metric, such as questionnaires or performance tests. Once an outcome has been defined, researchers typically want to know if any other factors can influence the result. This is typically assessed with regression analysis.

Regression analysis[1] determines the relationship of an independent variable (such as bone mineral density) on a dependent variable (such as ageing) with the statistical assumption that all other variables remain fixed. The calculation of the relationship results in a theoretical straight line, and the correlation co-efficient (r) measures how closely the observed data are to the theoretical straight line that we have calculated.

In such a linear model, we can judge how well the line fits the data ('goodness of fit') by calculating the coefficient of determination (or square of the regression line, R^2). R^2 is a measure of the percentage of total variation in the dependant variable that is accounted for by the independent variable. An R^2 of 1.0 indicates that the data perfectly fit the linear model. Any R^2 value less than 1.0 indicates that at least some variability in the data cannot be accounted for by the model (e.g., an R^2 of 0.5 indicates that 50% of the variability in the outcome data cannot be explained by the model).

Given these statistical tools, we can use the regression equation to predict the value of the dependent variable based on the known value of independent variable. Since many variables may contribute to the outcome (dependent variable), further statistical analysis can be achieved with multiple regression analysis. These models are essentially the same as simple regression analysis, except that the multiple regression analysis equation describes the interrelationship of many variables and allows us to evaluate the joint effect of these variables on the outcome variable in question.

Poitras et al[2] report an interesting study this month that aims to predict length of stay and early clinical function following joint arthroplasty. Multiple linear regression analyses produced an equation based on the timed-up-and-go test, which was associated with length of stay. In addition, models based on the pre-operative WOMAC function subscore produced the best model for describing early post-operative function (as calculated by the Older American Resources and Services ALD score). As such the authors were able to conclude that the outcomes assessments (timed-up-and-go and WOMAC) were predictive of outcome, and further modelling identified thresholds of the outcome assessment scores that related to better and worse outcomes.

How should we interpret these findings? The authors quite correctly suggest that models such as these could be of value in discharge planning and resource utilisation by targeting the patients that most need intervention and rehabilitation. The reported R^2 for the models, however, was 0.18. Bearing in mind that R^2, the coefficient of determination, measures the percentage of the variation in the dependent variable that is explained by variation in the independent variable,[3] taking the compliment ($100 - R^2$) we see that 82% of the variation in the outcome parameter assessed is unexplained by the model. The principal problem is that the variance in the population studied can strongly influence R^2 magnitude. Therefore, there is no guarantee that a high coefficient of determination is indicative of 'goodness of fit'. Similarly there is no guarantee that a small R^2 indicates a weak relationship, given that the statistic is largely influenced by variation in the independent variable.[4]

D. F. Hamilton, PhD, BSc (Hons), Research Fellow, Department of Trauma and Orthopaedics
University of Edinburgh, FU413 Chancellor's Building, 49 Little France Crescent, Edinburgh, EH16 4SB, UK.

M. Ghert, FRCS, FRCS (Ed), DM, Associate Professor, Department of Surgery
Deputy Editor, The Bone and Joint Journal, 22 Buckingham Street, London, WC2N 6ET, UK.

A. H. R. W. Simpson, FRCS, FRCS (Ed), DM, Editor-in-Chief The Bone and Joint Journal, 22 Buckingham Street, London, WC2N 6ET, UK.

Correspondence should be sent to Professor A.H. R. W. Simpson; e-mail: e.vodden@boneandjoint.org.uk

doi:10.1302/2046-3758.49.2000571 $2.00

Bone Joint Res 2015;4:152–153.

Therefore, there is no rule for interpreting the strength of R^2 in its application to clinical relevance. Useful high values of R^2 can be obtained with clinical data sets,[5] however, a low R^2 can still provide a useful clinical model with respect to data trends, but may be low in precision. In this study there is an association between the performance tests and length of stay; and, using the equations, we can indeed predict one from the other. The accuracy of this prediction though, needs to be borne in mind when using it as a clinical tool.

Furthermore, it is not rational to compare R^2 across different samples, which given clinical populations, are likely to differ significantly in the variance of the independent and dependent variables.[6]

In controlled environments, such as biomechanical tests on cadaveric bones, the variance across predictive measurements is likely to be low, and therefore R^2 values can be expected to lie in the 0.8 range.[7] In clinical studies, however, R^2 values vary widely depending on the nature of the analysis. For example, when comparing radiographic parameters or associating surgical technical factors, values of R^2 are reported in the 0.2 to 0.4 range.[8,9] Whereas, comparing data between separate (but intrinsically similar) outcome assessment questionnaires can yield higher values in excess of 0.7.[10]

As such, further validation of the Poitras study[2] using new datasets and, ideally, confirmatory analysis of the findings using a much larger sample size, would be required before their regression model could be recommended for use clinically. This does not devalue the appropriateness – or indeed 'worthiness' – of reporting these findings in the literature, as the important clinical tools typically start as ideas in small datasets. As with all research papers, the reader requires a basic understanding of methodology to evaluate how relevant the results are to wider practice.

References

1. **Draper NR, Smith H.** Applied regression analysis. Wiley-Interscience, 1998.
2. **Poitras S, Wood KS, Savard J, Dervin GF, Beaule PE.** Predicting early clinical function after hip or knee arthroplasty *Bone Joint Res* 2015;4:145–151.
3. **SchroederLD, Sjoquist DL, Stephen PE.** Understanding regression analysis: an introductory guide. 1986, Sage Publications; Beverly Hills, California.
4. **Filho DBF, Silva JA, Rocha E.** What is R2 all about? *Leviathan – Cadernos de Pesquisa Política* 2011;3:60–68.
5. **Maempel JF, Clement ND, Brenkel IJ, Walmsley PJ.** Validation of a prediction model that allows direct comparison of the Oxford Knee Score and American Knee Society clinical rating system. *Bone Joint J* 2015;97-B:503–509.
6. **Kennedy P.** A guide to econometrics. 2008, Wiley-Blackwell; San Francisco, California:27.
7. **Eckstein F, Wunderer C, Boehm H, et al.** Reproducibility and side differences of mechanical tests for determining the structural strength of the proximal femur. *J Bone Miner Res* 2004;19:379–385.
8. **Weber M, Lechler P, von Kunow F, et al.** The validity of a novel radiological method for measuring femoral stem version on anteroposterior radiographs of the hip after total hip arthroplasty. *Bone Joint J* 2015;97-B:306–311.
9. **Kuwashima U, Okazaki K, Tashiro Y, et al.** Correction of coronal alignment correlates with reconstruction of joint height in unicompartmental knee arthroplasty. *Bone Joint Res* 2015;4:128–133.
10. **Parsons N, Griffin XL, Achten J, Costa ML.** Outcome assessment after hip fracture: is EQ-5D the answer? *Bone Joint Res* 2014;19;3:69–75.